融合内容与矩阵分解的混合推荐算法 *

王永贵，陈玉伟†

(辽宁工程技术大学 软件学院，辽宁 葫芦岛 125105)

摘　要：传统的基于内容的推荐算法往往具有较低的准确性，而协同过滤推荐算法中普遍存在数据稀缺性和项目冷启动问题。为解决上述问题，提出了一种融合内容与协同矩阵分解技术的混合推荐算法。该算法实现了在共同的低维空间中分解内容和协同矩阵，同时保留数据的局部结构。在参数优化方面利用一种基于乘法更新规则的迭代方法，以此提高学习能力。实验结果表明，该算法优于其他具有代表性的项目冷启动推荐算法，有效缓解了数据稀疏性，提高了推荐准确性。

关键词：混合推荐；矩阵分解；冷启动；参数优化；局部结构

中图分类号：TP391　　doi: 10.19734/j.issn.1001-3695.2018.11.0797

Hybrid recommendation algorithm based on content and matrix factorization

Wang Yonggui, Chen Yuwei†

(*College of Software*, *Liaoning Technical University*, *Huludao Liaoning* 125105, *China*)

Abstract: Traditional content-based recommendation algorithms have lower accuracy, while data sparseness and cold start problems are common in collaborative filtering recommendation algorithms. To solve this problem, this paper proposed a hybrid recommendation algorithm based on content and collaborative matrix factorization technique. The algorithm realized the decomposition of content and collaborative matrix in a common low-dimensional space while preserving the local data structure. This paper used an iterative method based on multiplication update rules in parameter optimization, improved learning ability. The experimental results show that the proposed algorithm is superior to other representative projects cold start recommendation algorithm, which effectively alleviates the data sparseness and improves the efficiency of the algorithm.

Key words: hybrid recommendation; matrix factorization; cold start; parameter optimization; local structure

0　引言

随着大数据时代的到来，如何更好地处理运用好用户的信息数据，解决信息过载现象，已经成为互联网领域重要的课题之一。于是个性化推荐系统应运而生[1]。目前推荐算法已得到非常广泛的应用，如亚马逊、淘宝、网易云音乐等都具有强大的推荐系统，因其能根据海量的用户历史信息进行筛选过滤，挖掘出深层次的用户与用户或项目之间的关系，产生较为准确的具有偏好特性的个性化推荐，可以更好地满足用户需求，深受用户喜爱。

推荐算法主要可以分为基于内容的推荐[2]、协同过滤推荐[3]以及混合推荐算法[4]。基于内容的推荐算法通过抽取项目属性进行特征学习，根据计算出的项目相似度向用户推荐相似度值较大的项目。协同过滤算法是目前应用研究最为广泛的算法，主要可分为基于用户的协同过滤推荐和基于项目的协同过滤推荐，两者的核心在于通过用户—项目评分矩阵计算出用户或项目之间的相似度并进行推荐。但在现实应用中，只有一小部分用户会对少部分项目评分或评论，因此协同过滤技术存在明显的数据稀疏问题和项目冷启动问题[5]。冷启动问题主要可分为用户冷启动和物品冷启动，即如何为新用户推荐物品，或如何为用户推荐新物品。基于内容的方法虽然可以缓解冷启动问题，但是自身的推荐准确率较低，很少成为唯一选择。

为解决上述问题，本文将基于内容的推荐算法与矩阵分解结合，提出了一种混合推荐算法。其主要贡献可概括如下：

a）将内容与协同信息结合在统一的矩阵分解框架中，同时利用数据的局部结构，通过最近邻居形成的权重矩阵测量低维表示的局部平滑度。

b）利用一种基于乘法更新规则的学习模型进行参数学习，同时施加非负性约束，引入正则化，优化了目标函数，减少了运行时间。

c）在 nips12raw_str602 公开数据集上进行了大量的实验，证明了本文算法优于其他几种项目冷启动推荐算法，提高了推荐准确率，有效解决了项目冷启动问题。

1　相关工作

混合推荐算法的主要手段是综合利用基于内容的推荐算法和协同过滤推荐算法的优点，并避免各自的不足之处，可以有效缓解冷启动问题，实现多样性；但往往会出现复杂度过高、推荐时间增长、不易实现平衡等问题。常见的混合推荐算法主要可以分为加权型[6]、合并型[7]和特征组合[8]等几种类型。加权型混合推荐是将多种推荐计算结果加权后输出最终推荐结果。合并型混合推荐利用多种推荐算法产生多种推荐结果。特征组合混合推荐则是将不同推荐数据源的特征结合起来产生推荐。下面简要介绍几种相关的推荐算法。

Soboroff 等人[9]提出了一种基于潜在语义索引的技术，

收稿日期：2018-11-01；**修回日期：**2018-12-27　**基金项目：**国家自然科学基金面上项目（61772249）

作者简介：王永贵（1967-），男，内蒙古宁城人，教授，硕导，硕士，主要研究方向为大数据分析、云计算、机器学习；陈玉伟（1995-），男，辽宁营口人，硕士研究生，主要研究方向为机器学习、推荐系统(2584029560@qq.com).

是较早结合协同过滤和文档内容进行文本推荐的混合推荐算法。该方法通过为每个用户构建内容配置文件来发现集合中的主题，然后将与用户配置文件最相似的文档推荐给用户。Singh 等人[10]提出了协同矩阵分解的方法，是多关系分解模型的一个通用框架，解决了项目推荐和评分预测问题。Gantner 等人[11]使用 KNN（K-nearest-neighbor）和 BPR（Bayesian personalized ranking）优化学习项目或用户与潜在因子间的映射。Wang 等人[12]提出一个混合推荐框架，利用用户上下文、操作记录和用户特征将用户分类，然后根据分类动态选择推荐算法。Ahn[13]引入了一种混合相似度计算方法，融入了个性化信息，缓和了冷启动问题，但算法实现复杂度过高。Rosen-Zvi 等人[14],提出了一种生成概率模型 ATM（author-topic model），将每个作者与相关主题的多项分布关联，每个主题都具有关于词汇的多项分布。在新闻推荐中，将用户建模为作者。Felfernig 等人[15]提出了将高维评分矩阵映射到低维特征矩阵，利用用户的隐性数据和社交信息进行推荐，提高了推荐准确度，但存在部分信息丢失的问题。Cai 等人[16]利用了数据局部几何结构更好地进行低维表示。该算法受最近邻图的标签传播启发，提出了一种施加约束的聚类技术，取得了较好的效果。

以上算法融入了不同的因素，均有效缓解了冷启动问题，但存在复杂度过高、信息缺失等问题。本文算法融入了上述算法的部分优点，同时避免了信息缺失问题，利用了矩阵低秩特性和局部特性，并施加非负性约束来加强效应，采用基于乘法更新规则的方法优化参数，有效解决了冷启动问题，提高了推荐准确性。

2　融合内容和矩阵分解的混合推荐算法

对于初始项目矩阵 $\boldsymbol{X_I}$ 和用户矩阵 $\boldsymbol{X_U}$，利用参数矩阵 $\boldsymbol{W}$、$\boldsymbol{H_I}$ 和 $\boldsymbol{H_U}$ 对两者进行协同分解，加入局部特性对分解后的数据进行处理优化，定义出由参数 α、β、λ 控制的优化目标函数，利用一种参数学习算法，使参数矩阵 $\boldsymbol{W}$、$\boldsymbol{H_I}$ 和 $\boldsymbol{H_U}$ 在优化目标函数的背景下进行迭代学习，最终达到一个平稳点，产生推荐结果。

2.1　问题描述

针对项目冷启动问题，应向潜在感兴趣的用户推荐新项目，即至今没表达出兴趣但可能感兴趣的项目。根据用户的历史记录和新项目的描述，检索出最有可能对此新项目感兴趣的用户。可以将此问题定义如下：对于一个具有 n 个项目的集合，存储在矩阵 $\boldsymbol{X_I}\in \boldsymbol{R}^{n\times m}$ 中的 m 表示项目属性个数，矩阵的一行表示一个项目，矩阵的一列表示一个项目属性，见表 1；存储在矩阵 $\boldsymbol{X_U}\in \boldsymbol{R}^{n\times u}$ 中的 u 表示用户数量，其中一个单元（i，j）表示用户 j 是否对项目 i 有兴趣，见表 2。

基于以上定义，每个项目及对应的描述与所消费的用户相关联，如新闻。每篇新闻都由词汇和产生评论的用户来描述。这种情况可以用两个矩阵表示出来，其中一个矩阵表示为项目矩阵 $\boldsymbol{X_I}\in \boldsymbol{R}^{n\times v}$，$n$ 为文档数目，v 表示词汇量；另一个矩阵表示为用户矩阵 $\boldsymbol{X_U}\in \boldsymbol{R}^{n\times u}$，$n$ 为文档数目，u 表示用户数量。项目矩阵 $\boldsymbol{X_I}$ 可以用来表示文档中词汇的 TF-IDF 分数，用户矩阵 $\boldsymbol{X_U}$ 用来反映相关用户是否对当前文档产生评论。

将 $\boldsymbol{X_I}$ 分解成两个低维矩阵可以发现文档主题以及每个文档所属主题的程度；同样的，将 $\boldsymbol{X_U}$ 分解可以发现用户社区和文档在社区内引发兴趣的程度。这样分解的问题在于没有将文档和用户在共同的潜在空间表现出来，每个分解表示不同的潜在空间，无法将主题与社区联系起来。为使文档和用户在同一空间表示出来，使每个项目既能用一个主题描述，也能由一组用户即社区描述，本文将 $\boldsymbol{X_I}$ 和 $\boldsymbol{X_U}$ 分解后在共同的空间进行低维表示。对于矩阵 $\boldsymbol{X_I}$ 和 $\boldsymbol{X_U}$，定义如下的优化函数：

$$
\begin{aligned}
\min : Z = {} & \alpha \| \boldsymbol{X}_{\mathrm{I}} - \boldsymbol{W}\boldsymbol{H}_{\mathrm{I}} \|_{\mathrm{F}}^{2} + (1-\alpha) \| \boldsymbol{X}_{U} - \boldsymbol{W}\boldsymbol{H}_{\mathrm{U}} \|_{\mathrm{F}}^{2} \\
& + \lambda \left(\| \boldsymbol{W} \|_{\mathrm{F}}^{2} + \| \boldsymbol{H}_{\mathrm{I}} \|_{\mathrm{F}}^{2} + \| \boldsymbol{H}_{\mathrm{U}} \|_{\mathrm{F}}^{2} \right) \\
& s.t. \boldsymbol{W} \ge 0, \boldsymbol{H}_{\mathrm{I}} \ge 0, \boldsymbol{H}_{\mathrm{U}} \ge 0
\end{aligned}
\tag{1}
$$

其中：前两项是通过使用同一矩阵 $\boldsymbol{W}$ 实现在同一潜在空间下的项目矩阵 $\boldsymbol{X_I}$ 和用户矩阵 $\boldsymbol{X_U}$ 的分解，其中的符号表示矩阵的 F-范数，运算后可实现对 $\boldsymbol{X_I}$ 和 $\boldsymbol{X_U}$ 的降维表示。其中超参数 α 用来控制矩阵分解的重要程度，控制在[0,1]内。若 α=0.5，则表示两个分解同等重要；若 α<0.5，则表明用户矩阵 $\boldsymbol{X_U}$ 的分解更重要；若 α>0.5，则表明项目内容矩阵 $\boldsymbol{X_I}$ 的分解更重要。另一项为由超参数 $\lambda\geqslant 0$ 控制的关于参数矩阵 $\boldsymbol{W}$、$\boldsymbol{H_I}$ 和 $\boldsymbol{H_U}$ 的正则化项，用来防止过拟合，衡量函数光滑的程度。

表 1　项目—属性表

Table 1　Item-attribute

项目	属性			
	ATTR_1	ATTR_2	…	ATTR_m
Item_1	$\boldsymbol{R}_{11}$	$\boldsymbol{R}_{12}$	…	$\boldsymbol{R}_{1m}$
Item_2	$\boldsymbol{R}_{21}$	$\boldsymbol{R}_{22}$	…	$\boldsymbol{R}_{2m}$
⋮	⋮	⋮		⋮
Item_n	$\boldsymbol{R}_{n1}$	$\boldsymbol{R}_{n2}$	…	$\boldsymbol{R}_{nm}$

表 2　项目—用户表

Table 2　Item-user

项目	用户			
	User_1	User_2	…	User_u
Item_1	$\boldsymbol{R}_{11}$	$\boldsymbol{R}_{12}$	…	$\boldsymbol{R}_{1u}$
Item_2	$\boldsymbol{R}_{21}$	$\boldsymbol{R}_{22}$	…	$\boldsymbol{R}_{2u}$
⋮	⋮	⋮		⋮
Item_n	$\boldsymbol{R}_{n1}$	$\boldsymbol{R}_{n2}$	…	$\boldsymbol{R}_{nu}$

2.2　局部结构的特性与衡量

当进行协同分解时，应找到一个共同的低维空间对不同维数的数据进行处理优化，如式（1）所示。在考虑局部结构时引入流型假设，这种假设主要考虑模型的局部特性，如果两个数据点 x_i 和 x_j 在原有分布几何中距离接近，那么在低维空间中也应该彼此接近，这在降维[17]和半监督学习[18]中有重要作用。但这一理论在分布的几何结构未知的情况下不能直接使用，然而根据流形学习[19]的研究,局部几何结构可以通过最近邻图的方式来建模。

假设一个图具有 n 个节点，每个节点表示一个数据点，找到每个点拥有的 g 个最近邻居，并将这些点连接起来。对这些连接边进行描述，可以采用二进制和加权两种表示方法。若采用二进制表示方法，则描述成最近邻居为 1，其余为 0；若采用加权方式，则可以用余弦相似度值描述。本文采用余弦相似度值进行加权表示，并构成邻接矩阵 $\boldsymbol{A}$，以此来测量两个数据点之间的局部紧密度。若 x_i=（x_1，y_1），x_j=（x_2，y_2），则 x_i 与 x_j 之间的余弦相似度计算如下：

$$
sim(x_i, x_j) = \frac{x_1 x_2 + y_1 y_2}{\sqrt{x_1^2 + y_1^2} \times \sqrt{x_2^2 + y_2^2}} \tag{2}
$$

运用矩阵 $\boldsymbol{W}$ 协同分解后，每个数据点 x_i 被低维映射成了 w_i。测量两个低维数据点间的距离是通过计算欧几里德距离，即$\| \omega_i - \omega_j \|^2$。利用邻接权重矩阵 $\boldsymbol{A}$ 可以测量低维表示的局部平滑度。给定损失函数如下：

$$
\begin{aligned}
S &= \frac{1}{2}\sum_{i,j=1}^{n} \| \omega_i - \omega_j \|^2 \boldsymbol{A}_{ij} = \sum_{i=1}^{n} (\omega_i^T \omega_i) \boldsymbol{D}_{ii} - \sum_{i,j=1}^{n} (\omega_i^T \omega_j) \boldsymbol{A}_{ij} \\
&= \mathrm{Tr}(\boldsymbol{W}^T \boldsymbol{D} \boldsymbol{W}) - \mathrm{Tr}(\boldsymbol{W}^T \boldsymbol{A} \boldsymbol{W}) = \mathrm{Tr}(\boldsymbol{W}^T \boldsymbol{L} \boldsymbol{W})
\end{aligned}
\tag{3}
$$

其中：$\boldsymbol{D}$ 为对角矩阵；其对角值为 $\boldsymbol{A}$ 中每行元素的和；$D_{ii}=\sum_j A_{ij}$。Tr 表示矩阵的迹；$\boldsymbol{L}$ 为拉普拉斯矩阵[20]。

为了融入这种局部特性，针对式（1）进行优化，加入参数 β 来控制局部性的执行程度。优化后的目标函数如下：

$$\begin{gathered}\min: Z=\alpha\|\boldsymbol{X}_I-\boldsymbol{W}\boldsymbol{H}_I\|_F^2+(1-\alpha)\|\boldsymbol{X}_U-\boldsymbol{W}\boldsymbol{H}_U\|_F^2+\\ \beta\mathrm{Tr}(\boldsymbol{W}^T\boldsymbol{L}\boldsymbol{W})+\lambda\left(\|\boldsymbol{W}\|_F^2+\|\boldsymbol{H}_I\|_F^2+\|\boldsymbol{H}_U\|_F^2\right)\\ s.t.\ \boldsymbol{W}\ge 0, \boldsymbol{H}_I\ge 0, \boldsymbol{H}_U\ge 0\end{gathered} \tag{4}$$

其中：α 控制 $\boldsymbol{X}_I$ 和 $\boldsymbol{X}_U$ 分解的重要程度；λ 控制正则化，防止过拟合；$\boldsymbol{L}$ 为式（3）中的拉普拉斯矩阵。

2.3 基于乘法更新的参数学习算法

上述目标函数是想通过迭代来寻找最小值，但由于参数 $\boldsymbol{W}$、$\boldsymbol{H}_I$、$\boldsymbol{H}_U$ 都是非凸的，所以找到全局最小值非常困难。为此推导出一种基于乘法更新规则的迭代算法，进行参数学习，可以实现一个平稳点。首先求出关于 $\boldsymbol{W}$、$\boldsymbol{H}_I$、$\boldsymbol{H}_U$ 的偏导数，如式（5）～（7）所示。

$$\begin{gathered}\nabla_{\boldsymbol{W}}Z=\alpha\boldsymbol{W}\boldsymbol{H}_I\boldsymbol{H}_I^T-\alpha\boldsymbol{X}_I\boldsymbol{H}_I^T+(1-\alpha)\boldsymbol{W}\boldsymbol{H}_U\boldsymbol{H}_U^T-\\ (1-\alpha)\boldsymbol{X}_U\boldsymbol{H}_U^T+\beta\boldsymbol{L}\boldsymbol{W}+\lambda\boldsymbol{W}\end{gathered} \tag{5}$$

$$\nabla_{\boldsymbol{H}_I}Z=\alpha\boldsymbol{W}^T\boldsymbol{W}\boldsymbol{H}_I-\alpha\boldsymbol{W}^T\boldsymbol{X}_I+\lambda\boldsymbol{H}_I \tag{6}$$

$$\nabla_{\boldsymbol{H}_U}Z=(1-\alpha)\boldsymbol{W}^T\boldsymbol{W}\boldsymbol{H}_U-(1-\alpha)\boldsymbol{W}^T\boldsymbol{X}_U+\lambda\boldsymbol{H}_U \tag{7}$$

然后再利用 KKT（karush-kuhn-tucker）一阶最优性条[21]，可以推导出如下结论：

$$\boldsymbol{W}\ge 0, \boldsymbol{H}_I\ge 0, \boldsymbol{H}_U\ge 0 \tag{8}$$

$$\nabla_{\boldsymbol{W}}Z\ge 0, \nabla_{\boldsymbol{H}_I}Z\ge 0, \nabla_{\boldsymbol{H}_U}Z\ge 0 \tag{9}$$

$$\boldsymbol{W}\odot\nabla_{\boldsymbol{W}}Z=0, \boldsymbol{H}_I\odot\nabla_{\boldsymbol{H}_I}Z=0, \boldsymbol{H}_U\odot\nabla_{\boldsymbol{H}_U}Z=0 \tag{10}$$

其中：⊙ 为矩阵乘法运算符。再将式（5）～（7）分别带入式（10）中的三个结论，可以得到参数 $\boldsymbol{W}$、$\boldsymbol{H}_I$、$\boldsymbol{H}_U$ 的更新规则：

$$\boldsymbol{W}\leftarrow\boldsymbol{W}\odot\frac{[\alpha\boldsymbol{X}_I\boldsymbol{H}_I^T+(1-\alpha)\boldsymbol{X}_U\boldsymbol{H}_U^T+\beta\boldsymbol{A}\boldsymbol{W}]}{[\alpha\boldsymbol{W}\boldsymbol{H}_I\boldsymbol{H}_I^T+(1-\alpha)\boldsymbol{W}\boldsymbol{H}_U\boldsymbol{H}_U^T+\beta\boldsymbol{D}\boldsymbol{W}+\lambda\boldsymbol{W}]} \tag{11}$$

$$\boldsymbol{H}_I\leftarrow\boldsymbol{H}_I\odot\frac{[\alpha\boldsymbol{W}^T\boldsymbol{X}_I]}{[\alpha\boldsymbol{W}^T\boldsymbol{W}\boldsymbol{H}_I+\lambda\boldsymbol{H}_I]} \tag{12}$$

$$\boldsymbol{H}_U\leftarrow\boldsymbol{H}_U\odot\frac{\left[(1-\alpha)\boldsymbol{W}^T\boldsymbol{X}_U\right]}{\left[(1-\alpha)\boldsymbol{W}^T\boldsymbol{W}\boldsymbol{H}_U+\lambda\boldsymbol{H}_U\right]} \tag{13}$$

基于上式，$\boldsymbol{W}$、$\boldsymbol{H}_I$、$\boldsymbol{H}_U$ 通过迭代可以达到学习的效果，最终实现一个平稳点。目标函数 Z 在 $\boldsymbol{W}$、$\boldsymbol{H}_I$、$\boldsymbol{H}_U$ 处于稳定点时的更新规则下是不变的。本文又考虑到项目矩阵 $\boldsymbol{X}_I$ 和用户矩阵 $\boldsymbol{X}_U$ 的样本个数，将样本数的影响加入目标函数形成最终的优化目标函数：

$$\begin{gathered}\min: Z=\frac{\alpha}{\lg(v_1)}\|\boldsymbol{X}_I-\boldsymbol{W}\boldsymbol{H}_I\|_F^2+\frac{(1-\alpha)}{\lg(v_2)}\|\boldsymbol{X}_U-\boldsymbol{W}\boldsymbol{H}_U\|_F^2\\ +\beta\mathrm{Tr}(\boldsymbol{W}^T\boldsymbol{L}\boldsymbol{W})+\lambda(\|\boldsymbol{W}\|_F^2+\|\boldsymbol{H}_I\|_F^2+\|\boldsymbol{H}_U\|_F^2)\\ s.t.\ \boldsymbol{W}\ge 0, \boldsymbol{H}_I\ge 0, \boldsymbol{H}_U\ge 0\end{gathered} \tag{14}$$

其中：v_1 表示项目数；v_2 表示用户数。该优化目标函数能减少迭代次数，有效减少运行时间。一旦模型训练学习 $\boldsymbol{W}$、$\boldsymbol{H}_I$、$\boldsymbol{H}_U$，这些参数就具有了预测作用。例如给定一篇新的新闻文章描述词就可以预测最有可能评论这篇文章的用户 q_i。通过使用最小二乘法[22]求解出 $q_i=w\boldsymbol{H}_i$，然后将文档向量 q_i 投影到公共潜在空间。不断更新的 w 会在潜在空间中捕获到 q_i，并计算出 $w\boldsymbol{H}_u$，即用户评论新文章的可能性 q_u，从而根据这些可能性评分对这些用户进行排名。

3 实验结果与分析

3.1 实验环境

本文实验采用的硬件环境：AMD A8-5545M APU 四核，主频 1.70 GHz，内存 4 GB，硬盘 750 GB；操作系统：Windows7 操作系统；编程环境：MATLAB R2016a，Microsoft VC＋＋ 2015。

3.2 数据集描述

实验采用的是 nips12raw_str602 公开数据集，可以在 https://cs.nyu.edu/~roweis/data.html 找到。该数据集为稀疏数据集，对文章内容进行了预处理，去除停用词、数字、标点以及不常用标记，包含 2 037 篇文章和用户名、13 649 个描述词汇以及 1 740 个标题。

3.3 实验设置

3.3.1 实验方法

按时间顺序对数据集进行排序，并通过移动时间窗生成训练集和测试集，其中还将测试集限制为至少评论过一次文章的用户。然后采用 k 折交叉验证[23]，将数据集平分成 10 组，每次选取其中一组数据作为测试集，剩余组数据作为训练集。每个实验进行 10 次，取平均值作为最终的实验结果，并与其他算法比较分析。

3.3.2 评价指标

推荐系统常用的性能评价指标有准确率（precision）、召回率（recall）、平均绝对误差（mean absolute error， MAE）和均方根误差（root mean square error， RMSE）等[24]。但这些都属于准确性指标，具有不适合评估排序性能的局限性。

本实验采用的评价指标为均值平均精度(mean average precision， MAP) 和归一化折损累计增益（normalized discounted cumulative gain，NDCG）[25]。两者都考虑到了位置因素对实验结果的影响。MAP 较好地平衡了准确率和召回率，排名越靠前 MAP 值就越高；NDCG 可以评估整个测试集中的用户和推荐列表，进行归一化处理。上述两个指标的相关定义如下：

a）MAP 的具体定义见式（15）（16）所示。

$$AP=\frac{1}{R}\times\sum_{r=1}^{R}\frac{r}{position(r)} \tag{15}$$

$$MAP=\sum_{q=1}^{Q}\frac{AP(q)}{Q} \tag{16}$$

其中：AP 表示平均准确率（average precision， AP）；R 表示文档个数；position（r）是第 r 篇文档在列表中的位置；Q 表示求出的 AP 值总数。

b）NDCG 是用 IDCG 进行归一化处理，表示 DCG(discounted cumulative gain)与 IDCG（ideal discounted cumulative gain）差的距离。相关定义如下：

$$DCG_p=\sum_{i=1}^{p}\frac{2^{rel_i}-1}{\log_2(i+1)} \tag{17}$$

$$IDCG_p=\sum_{i=1}^{|REL|}\frac{2^{rel_i}-1}{\log_2(i+1)} \tag{18}$$

$$NDCG_p=\frac{DCG_p}{IDCG_p} \tag{19}$$

式（17）是一种增加相关度影响比重的 DCG，表示前 p 个位置的文档产生的效益，其中 rel_i 表示第 i 个文档的相关度等级。IDCG 意为理想状态下的 DCG 最大值。｜REL｜表示前 p 个文档的相关性大小排序。

3.4 实验结果与分析

实验将本文算法的两种模型与其他四种推荐算法进行性

能比较。本文的两种模型为 CMF 和未构建最近邻图的 CMF，即 CMF（β=0）。用来比较的四种算法分别为较为经典的混合推荐算法 UP-LSI[9]、解决冷启动的推荐算法 BPR-KNN[11]、生成概率模型推荐算法 ATM[14]和一种较新的混合推荐算法 SGHR（semi-genetic hybrid recommendation）[26]，采用遗传算法与权重相结合的方式提高推荐性能。

实验从算法平均性能、潜在主题数量 *k* 的影响、内容对协同信息的权重 *α* 的影响、控制解的平滑度参数 *λ* 的影响、选取最近邻居数 *g* 的影响以及运行时间上对算法进行对比评估。

1）算法平均性能对比

图 1 显示的是本文算法的两个模型 CMF 和 CMF（*β*=0）的参数调至最佳时，在 10 个不同训练集和测试集下取得的平均性能。可以看出本文的两个模型在两个评价指标上都明显高于另外四种冷启动算法，MAP 值提升了 15%左右，NDCG 值提升了 5%~10%，并且模型 CMF 优于模型 CMF（*β*=0），证明了本文算法有效提升了推荐准确度。

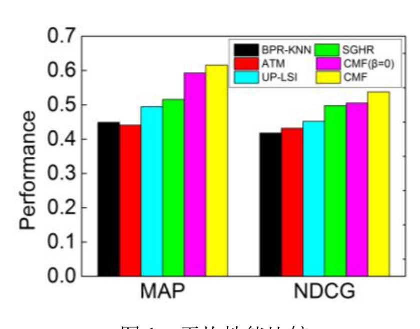


图 1 平均性能比较

Fig. 1 Average performance comparison

2）潜在主题数量的影响

主题数量 *k* 控制着模型的复杂度，若 *k* 值太小，即模型过于简单；若 *k* 值过大，会产生数据拟合，并且性能上会受到较大影响，导致性能较差。因此需找到合适的主题数量来平衡上述问题。本文算法将 *k* 控制在 100~1000。不同 k 值时的 NDCG 如图 2 所示。通过图 2 表明，主题数量在 400~700 时性能较稳定；当 *k*=600 时基本达到最佳性能；但当 *k*>700 后性能下滑明显，NDCG 值低于被对比算法；当 *k*>800 时模型 CMF（*β*=0）与 CMF 差距明显。但本文算法总体性能要远优于其他四种算法。

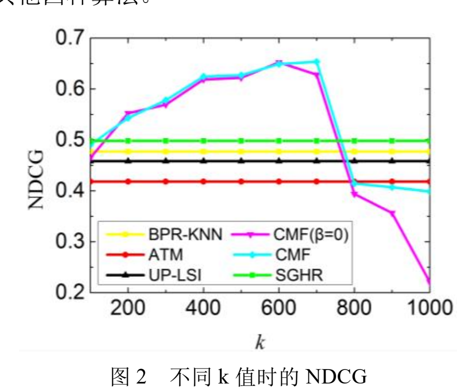


图 2 不同 k 值时的 NDCG

Fig. 2 NDCG at different k values

3）内容对协同信息的影响

参数 *α* 控制着内容对协同信息的重要程度。根据实验（2）结论，本实验设置潜在主题数量 *k*=600 来观察 *α* 的影响。其结果如图 3 所示。本文两种模型在 *α*∈[0.1,0.5]时性能较高较稳定，意味着协同分解时 X_U 更重要；当 *α*=0.2 时 CMF 达到最佳性能，且在 *α* 变化的整个过程中，CMF 整体性能优于 CMF（*β*=0），本文两种模型一直优于被对比算法。

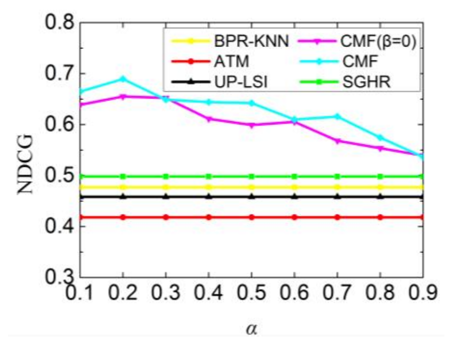


图 3 不同 α 值时的 NDCG

Fig. 3 NDCG at different α values

4）解的平滑度的影响

参数 λ 通过调整正则化项来控制解的平滑度。实验将 *k* 设定为 600，将 *α* 设定为 0.2，将控制局部平滑度的参数 *β* 设置为 0.25，同时控制 *λ*∈（0,1）。不同 λ 值时的 NDCG 如图 4 所示。由图 4 可以得出，当 *λ*∈[0.15,0.35]时可取得稳定高性能，CMF 性能略高于 CMF（β=0），同时两者远高于被对比算法；但当 *λ*>0.5 时性能较差，低于其他四种算法。

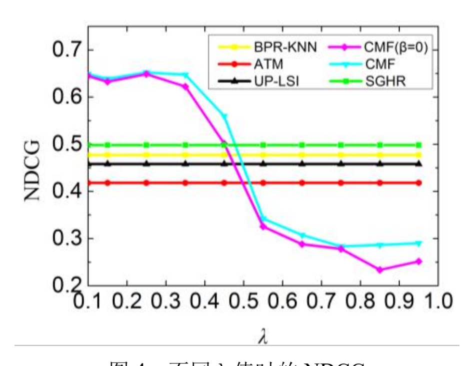


图 4 不同 λ 值时的 NDCG

Fig. 4 NDCG at different λ values

5）局部平滑度的影响

CMF 模型通过构建最近邻图形成的邻接权重矩阵 ***A*** 可计算局部平滑度，而参数 *β* 控制着局部平滑度的重要程度。图 5 显示了不同 *β* 值下的算法性能。当 *β*∈[0.1,0.9]时表现出较好的性能，且在 *β*=0.2 时 NDCG 值最高；当 *β*>1 时性能表现不佳。因此本实验将 *β* 控制在[0,1]内。

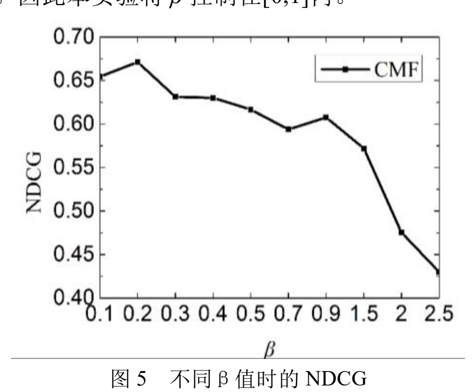


图 5 不同 β 值时的 NDCG

Fig. 5 NDCG at different β values

6）最近邻居数的影响

本文算法的 CMF 模型通过构建最近邻图来计算文档之间的相似度，而 CMF（*β*=0）模型并没采用这种方式，因此实验只针对 CMF 模型进行了最近邻居数影响的研究，如图 6 所示。可以看出，当最近邻居数在 1~4 个时，可以保持较高性能；当最近邻个数大于 5 时，则表现出较差的推荐准确性。

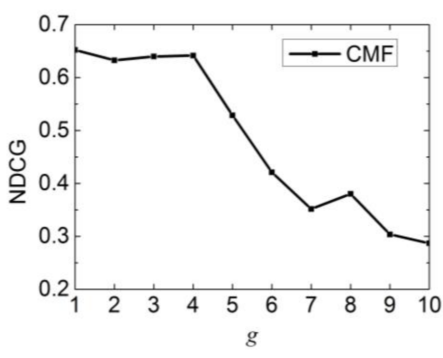


图 6　不同最近邻个数的 NDCG

Fig. 6　NDCG at different nearest neighbors

7）算法运行时间

将参数分别设置为 α=0.3，λ=0.25，β=0.25。测试本文两种模型在不同模型复杂度下的运行时间，在不同的 k 值下每个模型运行 10 次后取平均值，实验结果见图 7。可见 k>300 时 CMF 具有明显的优势，运行时间较 CMF（β=0）减少很多；在 k=800 时时间相差了 2 100 s 之多，可以得出加入最近邻加权和局部平滑度后的 CMF 模型可以有效减少运行时间；当 k>800 时由于实验性能较差，且运行时间较长，故不作考虑。

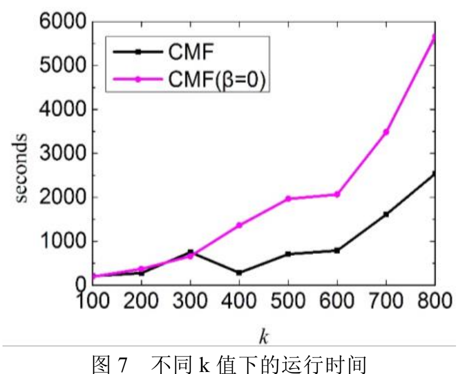


图 7　不同 k 值下的运行时间

Fig. 7　Run time at different k values

4　结束语

本文提出了内容与协同信息相结合的混合推荐算法。将内容与协同矩阵在共同的空间中分解，得到项目主题和主题程度矩阵，同时考虑到局部结构，通过构建最近邻图的方式计算用户间的偏好相似性，形成权重邻接矩阵，用来测量局部平滑度，施加了非负性约束，导致可解释和稀疏的潜在表示。

利用基于乘法更新规则进行参数学习，优化目标函数，找到性能稳定点。在评价指标上，使用了 MAP 和 NDCG，融入了位置因素，增加了相关度比重，可以更好地评估排序性能。本文算法进行了多次实验后与冷启动算法对比评估，结果显示本文算法显著优于对比算法，有效缓解了冷启动问题，提高了推荐准确性。并且通过本文两种模型之间的对比，证明了引入局部结构的有效性。

本文提出的融合内容和矩阵分解的算法虽然在该数据集上表现出较好的性能，但是否能支撑大规模数据集还有待研究。并且在今后的研究中，可以考虑加入更多的项目属性，如年龄、性别等属性，可以帮助提高用户的偏好相似性，从而提高推荐准确性。

参考文献：

[1] 陈洁敏，汤庸，李建国，等. 个性化推荐算法研究 [J]. 华南师范大学学报 :自然科学版 , 2014, 46 (5): 8-15. (Chen Jiemin, Tang Yong, Li Jianguo, *et al.* Survey of personalized recommendation algorithms [J]. Journal of South China Normal University :Natural Science Edition , 2014, 46 (5): 8-15.)

[2] 王立才，孟祥武，张玉洁. 上下文感知推荐系统 [J]. 软件学报, 2012, 23 (1): 1-20. (Wang Licai, Meng Xiangwu, Zhang Yujie. Context-aware recommender systems [J]. Journal of Software, 2012, 23 (1): 1-20.)

[3] Breese J S, Heckerman D, Kadie C. Empirical analysis of predictive algorithms forcollaborative filtering [C]// Proc of the 14th Conference on Uncertainty in Artificial Intelligence. San Francisco: Morgan Kaufmann Publishers Inc, 1998: 43-52.

[4] De Campos L M, Fernández-Luna J M, Huete J F, *et al.* Combining content-based and collaborative recommendations: a hybrid approach based on Bayesian networks [J]. International Journal of Approximate Reasoning, 2010, 51 (7): 785-799.

[5] 于洪，李俊华. 一种解决新项目冷启动问题的推荐算法 [J]. 软件学报, 2015, 26 (6): 1395-1408. (Yu Hong, Li Junhua. Algorithm to solve the cold-start problem in new item recommendations [J]. Journal of Software, 2015, 26 (6): 1395-1408.)

[6] 邹凌君，陈峻，李娟. 时间加权的混合推荐算法 [J]. 计算机科学, 2016, 43 (S2): 451-454. (Zou Lingjun, Chen Jun, Li Juan. Time-weighted hybrid recommender algorithm [J]. Computer Science, 2016, 43 (S2): 451-454.)

[7] 吴彦文，王洁. 基于混合遗传粒子群优化推荐算法的设计 [J]. 计算机工程与设计, 2017, 38 (2): 405-408. (Wu Yanwen, Wang Jie. Similarity recommendation based on hybrid of GA and PSO [J]. Computer Engineering and Design, 2017, 38 (2): 405-408.)

[8] 李克潮，蓝冬梅. 一种属性和评分的协同过滤混合推荐算法 [J]. 计算机技术与发展, 2013, 23 (7): 116-119,123. (Li Kechao, Lan Dongmei. A collaborative filtering hybrid recommendation algorithm for attribute and rating [J]. Computer Technology and Development, 2013, 23 (7): 116-119,123.)

[9] Soboroff I, Nicholas C. Combining content and collaboration in text filtering [C]// Proc of the IJCAI'99 Workshop on Machine Learning for Information Filtering. 1999: 86-91.

[10] Singh A , Gordon G. Relational learning via collective matrix factorization [C]// Proc of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, New York: ACM Press, 2008: 650-658.

[11] Gantner Z, Drumond L, Freudenthaler C, *et al.* Learning attribute-to-feature mappings for cold-start recommendations [C]// Proc of IEEE International Conference on Data Mining. Washington DC: IEEE Press, 2010: 176-185.

[12] Wang J H, Chen Yihao. A distributed hybrid framework to address the new-user cold-start problem [C]// Proc of the 13th IEEE International Conference on Ubiqutous Itelligence and Computing. Piscataway, NJ: IEEE Press, 2015: 1686-1691.

[13] Ahn H J. A new similarity measure for collaborative filtering to alleviate the new user cold-starting problem [J]. Information Sciences, 2008, 178 (1): 37-51.

[14] Rosen-Zvi M, Griffiths T, Steyvers M, *et al.* The author-topic model for authors and documents [C]// Proc of Uncertainty in Artificial Intelligence. 2004: 487-494.

[15] Felfernig A, Ninaus G, Grabner H, *et al.* An overview of recommender systems in requirements engineering [M]// Maalej W, Thurimella A K.

Managing Requirements Knowledge. Berlin: Springer, 2013: 315-332.

[16] Cai Deng, He Xiaofei, Wu Xiaoyun, *et al.* Non-negative matrix factorization on manifold [C]// Proc of the 8th IEEE International Conference on Data Mining. Pisa: IEEE Press, 2008: 63-72.

[17] Yin Jun, Lai Zhihui, Yan Hui. Dimension reduction using kernel collaborative representation based projection [J]. AEUE-International Journal of Electronics and Communications, 2017, 81 (11): 23-30.

[18] 刘建伟, 刘媛, 罗雄麟. 半监督学习方法 [J]. 计算机学报, 2015, 38 (8): 1592-1617. (Liu Jianwei, Liu Yuan, Luo Xionglin. Semi-supervised learning methods [J]. Chinese Journal of Computers, 2015, 38 (8): 1592-1617.)

[19] Allab K, Labiod L, Nadif M. Multi-manifold matrix decomposition for data co-clustering [J]. Pattern Recognition, 2017, 64 (4): 386-398.

[20] Katsnelson V. The matrix function\ (\varvec{e^{tA+B}\) is representable as the Laplace transform of a matrix measure [J]. Integral Equations and Operator Theory, 2016, 86 (3): 439-452.

[21] Cichocki A, Zdunek R, Phan A H, *et al.* Nonnegative matrix and tensor factorizations: applications to exploratory multi-way data analysis and blind source separation[M]. [S.l.]:Wiley, 2009.

[22] Dehghan M, Mohammadi V. Error analysis of method of lines (MOL) via generalized interpolating moving least squares (GIMLS) approximation [J]. Journal of Computational and Applied Mathematics, 2017, 321 (9): 540-554.

[23] Wang Meiling, Ma Jun. A novel recommendation approach based on users'weighted trust relations and the rating similarities [J]. Soft Computing, 2015, 20 (10): 3981-3990.

[24] Willmott C J , Matsuura K. Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance[J]. Climate Research, 2005,30:79-82.

[25] Keyhanipour A H, Moshiri B, Rahgozar M. CF-Rank: learning to rank by classifier fusion on click-through data [J]. Expert Systems With Applications, 2015, 42 (22) : 8597-8608.

[26] Mueller J. Combining aspects of genetic algorithms with weighted recommender hybridization [C]// Proc of the 19th International Conference on Information Integration and Web-based Applications & Services. New York: ACM Press, 2017: 13-22.